REMARKS

Claims 1-18 are pending in this application, of which claims 1, 17, and 18 are independent. Claim 1 has been amended to correct a minor typographical error. Claim 17 has been amended to address the 35 U.S.C. § 101 rejection that was raised by the Examiner in the Final Office Action mailed October 7, 2008 ("Final Action"). The amendments to the claims do not include new subject matter and do not require a new search. Favorable reconsideration of the Final Action is respectfully requested in view of the foregoing amendments and the following remarks.

35 U.S.C. § 101 Rejections

The preamble of claim 17 has been amended as suggested by the Examiner. Withdrawal of the 35 U.S.C. § 101 rejection of the claim is requested.

35 U.S.C. § 103 Rejections

Claims 1-15, 17, and 18 were rejected under 35 U.S.C. § 103(a) as being unpatentable over Chou et al. (US 5,797,123) in view of Foote ("An Overview of Audio Information Retrieval").

Claim 1 recites:[1]

> **1. A method comprising:**
>
> ***[A]* accepting first query data representing one or more spoken instances of a query in a first set of audio signals;**
>
> ***[B]* processing the first query data including determining a representation of the query that defines multiple sequences of subword units each representing the query;**
>
> ***[C]* accepting second speech data representing unknown speech in a second audio signal; and**
>
> ***[D]* locating putative instances of the query in the second speech data using the determined representation of the query.**

The "first query data" of limitation **[A]** represents, for example, a keyword that a user wishes to locate in an audio signal that represents unknown speech. Paragraph [030]

[1] Annotated with paragraph identifiers for ease of reference.

of the current application provides one example of "processing the first query data" recited in limitation **[B]**:

> The word spotting system 100 includes a query recognizer 150, which includes an implementation of a speech recognition algorithm and which is used to process acoustically-based data associated with the spoken query. The query recognizer 150 produces a processed query 160. The processed query 160 includes a data representation of the query in terms of subword linguistic units, which in this version of the system are English language phonemes. This representation of the query defines one or more possible sequences of subword units that can each correspond to the query. **The data representation of the processed query 160 defines a network representation of the query such that paths through the network each correspond to a possible sequence of subword units.**

Each of FIGS. 3 and 4 (reproduced below) of the current application depicts an exemplary representation of a query associated with the word "jury," where each representation defines multiple sequences of subword units (e.g., in FIG. 3, a first sequence of subword units is formed by y-uh-r-iy, a second sequence of subword units is formed by y-ih-r-iy, a third sequence of subword units is formed by y-uh-er-iy, etc.; in FIG. 4, a first sequence of subword units is formed by y-uh-r-iy, a second sequence of subword units is formed by y-er-iy; a third sequence of subword units is formed by jh-uh-r-iym, etc.).

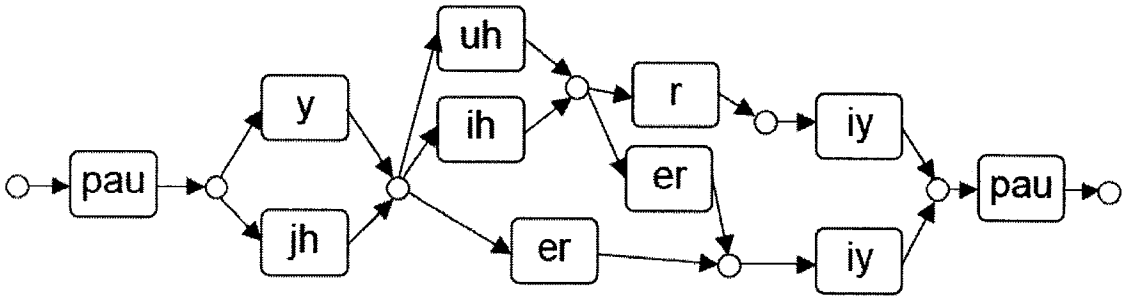

FIG. 3

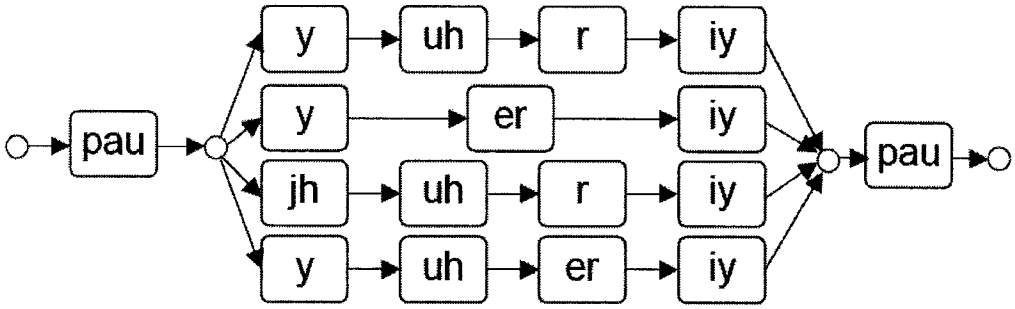

FIG. 4

The exemplary word spotting engine described in paragraph [031] of the current application uses a representation of the query associated with the word "jury" (e.g., representation depicted in FIG. 3 and/or FIG. 4) to process the unknown speech of limitation **[C]**, which is input to the word spotting system, to locate putative instances of the word "jury" in the unknown speech.

Next, we provide a brief description of the Chou reference.

In the Background section of the Chou reference (col. 1, lines 21-51), Chou states that one approach undertaken by prior art spoken dialogue recognition and understanding systems involves the use of deterministic finite state grammars (FSG), limited to the task or application at hand, to accept (and thereby to recognize and ultimately understand) user utterances. In such systems, the recognizer tries to match or decode the entire spoken input into any of the possible (i.e., acceptable in accordance with the fixed

grammar) word sequences. Chou recognizes that out-of-grammar utterances (e.g., extraneous words, hesitations, repetitions, and unexpected expressions) are typically encountered in most real world environments, and such out-of-grammar utterances reduce the effectiveness and/or performance of FSG-based spoken dialogue recognition and understanding systems in recognizing sub-tasks such as a spoken date or time.

In the Summary section of the Chou reference (col. 3, line 22 – col. 4, lines 10), Chou states that most spoken dialogue utterances (i.e., sentences) contain certain keywords and "key-phrases" that are task related, the recognition of which may advantageously lead to partial or full understanding of the utterance, while other portions of the utterance are not, in fact relevant to the task and thus should be ignored. To that end, Chou proposes a key-phrase detection and verification technique that applies a multiple pass procedure to one spoken utterance, so as to "partially or fully understand" that one spoken utterance. A sequence of four passes includes:

1. key-phrases are detected in that one spoken utterance;
2. the key-phrases are "verified" to reduce the set of detected key-phases to only those that exceed a confidence measure threshold;
3. sentence hypotheses are formed from the verified key-phrases;
4. sentence hypotheses are verified.

In rejecting claim 1, the Examiner points to different passes of the multiple pass sequence in support of his position that Chou teaches limitations **[A]** and **[B]** of claim 1. For ease of reference, the Examiner's comments (reproduced below in bolded text) with respect to specific limitations of claim 1 are followed by the Applicant's comments.

> **As per claim 1, Chou teaches the method comprising:**
> **accepting first query data representing one or more spoken instances of a query in a first set of audio signals; (Chou, columns 4-5, lines 65-67 and 1-9, ... In particular, this sentence hypothesis verification process is performed with a "partial input" comprising fewer subwords than are found in the entire utterance ... The input is an utterance which is a spoken instance of a query which is received.)**

The Examiner points to a portion of the Chou reference that describes sentence hypothesis verification (i.e., pass number **4** of the sequence) in support of his position that Chou teaches limitation **[A]** of claim 1. In particular, the Examiner reads the "'partial

input' comprising fewer subwords than are found in the entire utterance" that is accepted as input to the sentence verification process as corresponding to the recited "first query data representing one or more spoken instances of a query in a first set of audio signals" of claim 1.

> **processing the first query data including determining a representation of the query that defines multiple sequences of subword units each representing the query; (Chou, column 5, lines 60-65, ... The subword model recognizer employed by keyphrase detector 11 uses lexicon 23 and subword models 22, which may have been trained based, for example, on a conventional minimum classification error (MCE) criterion, familiar to those skilled in the art...)**

If we adopt the correspondence of elements set forth by the Examiner with respect to limitation **[A]** of the Chou reference, a consistent reading of limitation **[B]** of claim 1 would require that Chou disclose processing this "partial input," where the processing includes determining a representation of the "partial input" that defines multiple sequences of subword units each representing the query. Chou provides no such disclosure either in the portion of the Chou reference (col. 11, line 6 to col. 12, line 4) in which sentence hypothesis verification is described in detail or the portion of the Chou reference (col. 5, lines 60-65) cited by the Examiner in support of his position.

Recall that in the Summary section of the Chou reference, Chou identifies one limitation of prior art systems as being unable to recognize a sub-task such as a spoken date if the utterance is out-of-grammar. Chou's sentence hypothesis verification process is designed to overcome such a limitation by performing a semantic verification evaluation to determine whether a sentence hypothesis is semantically "legal" even in those instances in which the sentence hypothesis includes a "partial input" (aka an "incomplete utterance"; see col. 11, lines 41-43: "... a user might just say the month "August" without specifying any particular day of the month."). Chou is silent about processing the "partial input" (which the Examiner reads as corresponding to the recited "first query data") in any other context. Chou provides no hint or disclosure of processing the "partial input" where the processing includes determining a representation of the "partial input" that defines multiple sequences of subword units each representing the query.

The only context in which Chou provides any suggestion of a representation of a query that defines multiple sequences of subword units is with respect to the keyword detection unit. In col. 6, line 66 to col. 7, line 14, Chou states:

> ... the detection unit comprises a network of key-phrase sub-grammar automata with their permissible connections and/or iterations. Such automata can easily be extended to a stochastic language model by estimating the connection weights. The use of such models achieves wider coverage with only modest complexity when compared with sentence-level grammars. By way of illustration, FIG. 2 shows a simplified (i.e., reduced) phrase network example which may be used by key-phrase detector 11 of the illustrative system of FIG. 1 when applied to a "date retrieval" sub-task. A complete realization of this network example would allow virtually any iterations of days of the week, months, days of the month, and years, with certain appropriate constraints. (The total vocabulary size of such a complete realization is 99 words.) In this particular sub-task, no carrier phrases are incorporated.

Chou's "network of key-phrase sub-grammar automata" that is included in the detection unit does not result from processing a first query data that represents one or more spoken instances of a query in a first set of audio signals. Rather, Chou's "network of key-phrase sub-grammar automata" represents a predetermined "set of phrase subgrammars which may ... be specific to the state of the dialogue" (see col. 3, lines 44-52) that may be used by Chou's detection unit to detect the semantically significant portions of a sentence (see col. 3, lines 34-42) even if the sentence includes out-of-grammar utterances.

For the reasons given above, the applicant respectfully submits that Chou fails to disclose the features recited in limitation **[B]** of claim 1.

With respect to limitations **[C]** and **[D]** of claim 1, the Examiner states:

> **Chou fails to fully teach, but Foote teaches:**
> **accepting second speech data representing unknown speech in a second audio signal; and (Foote, abstract, ... This paper reviews the state of the art in audio information retrieval, and presents recent advances in automatic speech recognition, word spotting, speaker and music identification, and audio similarity with a view towards making audio less "opaque"... Audio information retrieval implies that there must be something to be retrieved from some audio signal. The second speech data is the source audio data to be searched.)**

locating putative instances of the query in the second speech data using the determined representation of the query. (Chou teaches the use of subword spotting for recognition. Foote further provides in sections 2.1 and 2.2 the use of keyword spotting with subunits for the purposes of information retrieval.)

It would have been obvious to someone of ordinary skill in the art at the time of the invention to combine Foote with the Chou device because all the claimed elements were known in the prior art and one skilled in the art could have combined the elements as claimed by known methods with no change in their respective functions, and the combination would have yielded predictable results to one of ordinary skill in the art at the time of the invention. Foote provides the use of subunit word spotting for information retrieval where Chou could further provide the method for word spotting as is recited in the claim language.

Foote provides an overview of a number of different audio information retrieval techniques. One commonality amongst the described techniques is the generation of a representation of the unknown speech that is to be searched. For example, in section 2.2, Foote describes a "lattice-based" word spotting technique that involves generating, by a phone or word recognition system, a lattice that is a compact representation of multiple best hypothesis of the unknown speech. This lattice may subsequently be searched to locate putative instances of a query.

However, claim 1 requires more than just locating putative instances of a query. Specifically, limitation **[D]** of claim 1 calls for "locating putative instances of the query ... using the determined representation of the query [that defines multiple sequences of subword units each representing the query]." Neither Chou nor Foote contemplate "determining a representation of the query that defines multiple sequences of subword units each representing the query" as recited in limitation **[B]** of claim 1. Accordingly, it is no surprise that neither reference provides any disclosure of using such a representation of a query to locate putative instances of the query within an unknown speech.

Chou and Foote, alone or in combination, provide no teaching or suggestion of the features of limitations **[B]** and **[D]** of claim 1. For at least the reasons stated above, claim 1 and its dependants are allowable over Chou and Foote. Should the Examiner choose to maintain the rejection of claim 1 as being unpatentable over Chou and Foote, the Examiner is respectfully requested to point out with specificity where in Chou and/or Foote the Examiner finds the alleged teaching of "determining a representation of the

query that defines multiple sequences of subword units each representing the query" as recited in limitation **[B]** of claim 1, and "locating putative instances of the query ... using the determined representation of the query [that defines multiple sequences of subword units each representing the query]" as recited in limitation **[D]** of claim 1.

Independent claims 17 and 18 contain similar limitations and are allowable over the cited references for at least the same reasons set forth above with respect to claim 1.

Conclusion

It is believed that all of the pending claims have been addressed. However, the absence of a reply to a specific rejection, issue or comment does not signify agreement with or concession of that rejection, issue or comment. In addition, because the arguments made above may not be exhaustive, there may be reasons for patentability of any or all pending claims (or other claims) that have not been expressed. Finally, nothing in this paper should be construed as an intent to concede any issue with regard to any claim, except as specifically stated in this paper, and the amendment of any claim does not necessarily signify concession of unpatentability of the claim prior to its amendment.

No fees are believed to be due in connection with filing of this application. However, please apply any charges or credits to Deposit Account No. 50-4189, referencing Attorney Docket No. 30004-004US1.

Respectfully submitted,

Date: 12/5/08

Mandy Jubang
Reg. No. 45,884

Customer No. 69713
Occhiuti Rohlicek & Tsao LLP
10 Fawcett Street
Cambridge, MA 02138
Telephone: (617) 500-2504
Facsimile: (617) 500-2499
59844.doc